7-21-2020

Application of genomic selection in commercial egg-type populations

J. E. Fulton
Hy-Line International

Anna Wolc
Iowa State University, awolc@iastate.edu

Application of genomic selection in commercial egg-type populations

Abstract

Genomic selection is a relatively new genetics tool that is being utilized by commercial poultry breeding companies to increase the rate of progress for performance traits. It utilizes information from multiple genetic variants (genotypes) located across the entire genome, in combination with trait information (phenotypes) for those individuals with genomic data. These two types of information are combined and used to predict the most accurate breeding value for individuals. The breeding value is the sum of the additive effects on phenotype of alleles at all single-nucleotide polymorphisms (SNP) and is twice the expected deviation from population mean in performance of the progeny when the individual is mated to other individuals from the same population. Thus, the genotype information can be used to predict which individual has the best genetics to pass to subsequent generations to improve progeny performance.

Disciplines

Agriculture | Animal Sciences | Genetics and Genomics | Poultry or Avian Science

Comments

This chapter is published as Fulton J.E., Wolc A. 2020. Application of genomic selection in commercial egg-type populations. In: Advances in Poultry Genetics and Genomics. Ed: Aggrey S.E., Zhou H., Tixier-Boichard M., Rhoads D.D., Burleigh Dodds Science Publishing, pp. 403-420. Posted with permission.

Chapter 15

Application of genomic selection in commercial egg-type populations

J. E. Fulton, Hy-Line International, USA; and A. Wolc, Hy-Line International and Iowa State University, USA

1 Introduction
2 Specific advantages of genomics for selection of egg production traits
3 Genomic selection versus phenotypic selection
4 Factors impacting genomic selection
5 Analysis methods for genomic selection
6 Conclusion
7 Future trends
8 Where to look for further information
9 References

1 Introduction

Genomic selection is a relatively new genetics tool that is being utilized by commercial poultry breeding companies to increase the rate of progress for performance traits. It utilizes information from multiple genetic variants (genotypes) located across the entire genome, in combination with trait information (phenotypes) for those individuals with genomic data. These two types of information are combined and used to predict the most accurate breeding value for individuals. The breeding value is the sum of the additive effects on phenotype of alleles at all single-nucleotide polymorphisms (SNP) and is twice the expected deviation from population mean in performance of the progeny when the individual is mated to other individuals from the same population. Thus, the genotype information can be used to predict which individual has the best genetics to pass to subsequent generations to improve progeny performance.

The concept of genomic selection was first proposed by Meuwissen et al. (2001). They suggested that utilization of information at the level of DNA in a selection program would result in more accurate identification of individuals with the desired genetics. This would improve the rate of genetic gain faster

http://dx.doi.org/10.19103/AS.2020.0065.21

than could be achieved by using phenotypic information only. More accurate identification of desired combinations of genotypes could then be used to select superior individuals for breeding and in subsequent generations increase the frequency of the desired variants. More accurate identification of superior individuals leads to improved rate of genetic gain in subsequent generations. Furthermore, they proposed that this genomic selection could be done at an early age, before phenotypic data would be available. This could shorten the generation interval, thus further increasing the rate of genetic progress.

The application of genomic selection could not be done until a large number of DNA markers were available and their genotypes could be obtained for a relatively inexpensive cost. The chicken was the first of the livestock species to have a genome sequence (Hillier et al., 2004). Simultaneously with the sequencing of the chicken genome, 2.8 million SNP were also identified from partial sequences from four chickens; one individual from an inbred strain of the White Leghorn breed, two commercial broilers, and one Chinese Silkie bird (Wong et al., 2004). This information provided the initial resources needed to develop large-scale and affordable DNA genotyping tools as it provided information on the existence of large numbers of SNP in chickens. From this information, SNP chips were developed, which allow the simultaneous detection of variation in multiple SNPs. The first SNP chip reported for the chicken contained 3072 SNPs and was used to examine SNP variation in chickens from both elite commercial production lines and experimental populations (Muir et al., 2008). Before large-scale SNP typing could be utilized by commercial breeders, sequence information from a wider source, including commercial layer and broiler lines, was needed. Kranis et al. (2013) produced genome sequences from 243 chickens, representing a diverse collection of 24 lines, including commercial egg production breeds used to produce either white or brown shell eggs, multiple broiler elite stocks, and inbred research lines. This encompassed far more of the genetic diversity currently found within the commercial chicken production industry than did the original four chickens sequenced. These sequences identified 139 million SNPs, a subset of which was then utilized to produce a 600K SNP Axiom chip that is now publicly available from Applied Biosystems.

2 Specific advantages of genomics for selection of egg production traits

Improvement for performance of the multiple traits involved with egg production is particularly challenging, as the majority of the traits under selection are expressed only in females and thus cannot be directly selected for in males. This includes the important traits of number of eggs laid, internal

and external egg-quality traits, and feed conversion (egg output vs feed intake). Furthermore, traits such as persistency of egg production, persistent shell quality, and livability can only be measured on hens after many months of production. Application of genomic selection allows for the identification of males with the desired genetic variants that can be passed on to their daughters, even though the males never lay eggs. It also allows for the early identification of hens with superior egg production, shell quality, and persistency traits, long before they lay eggs.

3 Genomic selection versus phenotypic selection

The Breeder's Equation was first introduced by Lush in 1943 (Lush, 1943) and states that the genetic gain per year depends on the intensity of selection, accuracy of the estimated breeding values, and the amount of genetic variation in the population, and is inversely proportional to the length of the generation interval. Any new technology that impacts any of these components can alter the rate of progress (Fig. 1). Theory suggests that the rate of improvement when utilizing genomic information is much greater than when only phenotypic selection is used because it affects each component of the Breeder's Equation.

3.1 Selection intensity

Traditional selection programs keep a large number of females since they can provide direct phenotypic information on the egg production traits of most interest. A smaller number of males are kept, with the majority of the males being randomly eliminated immediately at hatch, as all full-sib males have the same breeding values based on the phenotypes of their female relatives. With genomic selection, this process is changed and selection intensity can be increased by utilizing a two-stage selection scheme on the males. Initially, the number of male candidates that are hatched and reared can be increased. Genotype information can be quickly obtained and males with undesirable genotypes can be identified and then eliminated before they consume too many resources. The second stage of selection would occur once their sisters have phenotypes. Obtaining information on a larger number of males increases the selection intensity that can be placed on the males, thus improving progress on the various traits under selection.

3.2 Accuracy

As indicated previously, the most important traits for commercial egg production are expressed in females only (egg production and egg-quality-related traits). Thus, the accuracy of estimated breeding values of males prior

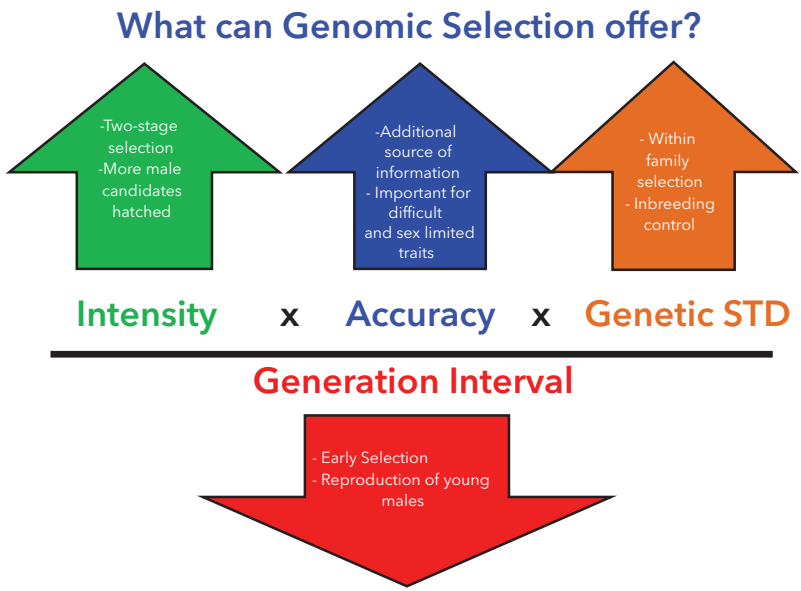

Figure 1 Impact of genomic selection on the components of the Breeder's Equation.

to availability of genomic information was very limited, unless progeny testing was used which subsequently increased generation intervals. Moreover, the phenotypes of the female relatives provide information only on which families are superior but do not differentiate breeding values for males within a family. Accuracy of estimated breeding values in these sex-limited traits can be greatly improved by genomic selection, as it utilizes the SNP genotypes of individual males, even though they do not express these traits. The gains in accuracy of estimated breeding values from using genomic information have been shown for a range of egg production and quality traits (Wolc et al., 2011b, 2013b), feed efficiency (Wolc et al., 2013a), and survival (Alemu et al., 2016; Brinker et al., 2016, 2018).

3.3 Genetic variation

Although intensity of selection can be increased by decreasing the number of individuals selected as parents of the next generation, this can also have the undesirable consequence of increasing inbreeding within the population. This results in a decrease in the genetic variation available for selection. Genetic selection using only pedigree information tends to favor a limited number of families, thus increasing inbreeding. By using genomic selection, which enables within-family selection, individuals with identical family information, such as full-sibs, can be further differentiated, thereby identifying favorable

individuals from more families. Furthermore, genomic information reduces the constraints on population structure. Traditional mating schemes in poultry have been hierarchical, with a group of females preassigned to a specific male. With the use of artificial insemination and hens in individual cages, chicks can be identified as to both sire and dam. Full pedigree information cannot be easily assigned in floor pens with natural mating with multiple males and females. With genomic information on the chicks and parents, full pedigree information can be accurately determined. For poultry, this means that females can be mated to multiple sires to increase the overall genetic variation of the progeny of each dam, while still maintaining full pedigree information.

3.4 Generation interval

Generation interval is defined as the average age of the parents at the birth of their offspring. Decreasing the generation interval allows annual genetic progress to increase more quickly. The generation interval for chickens is relatively short when compared with other livestock species and with layer chickens it is common practice to reproduce annually. With genomic information, males can be selected at a much earlier age, as there is no benefit to keeping them to an older age since they do not lay eggs. They can also reproduce at 6 months of age. In contrast, although females lay eggs at 6 months of age, the performance information (such as egg production and egg quality) requires many months to accumulate. Genomic information on the males allows identification of genetically superior younger males, who can then be mated to older hens which have their own phenotypic data. This allows the generation interval to be decreased in half for the sires, for an overall generation interval of 9 months compared to the standard 12-month generation interval for layer chickens.

4 Factors impacting genomic selection

4.1 Training panels

Before genomic selection can be implemented, a training population (or reference population), consisting of a set of individuals with both genotype and phenotype data, is required. This training panel enables the estimation of the effect of the genotyped SNPs on the specific traits of interest. This is the basic information that must be gathered before genomic selection can be implemented. A large training population that is closely related to the target population is needed to provide the most accurate genomic breeding values (Pszczola et al., 2012). The training panel should consist of at least 1000 individuals with both phenotypic and genotypic information (Oldenbroek and

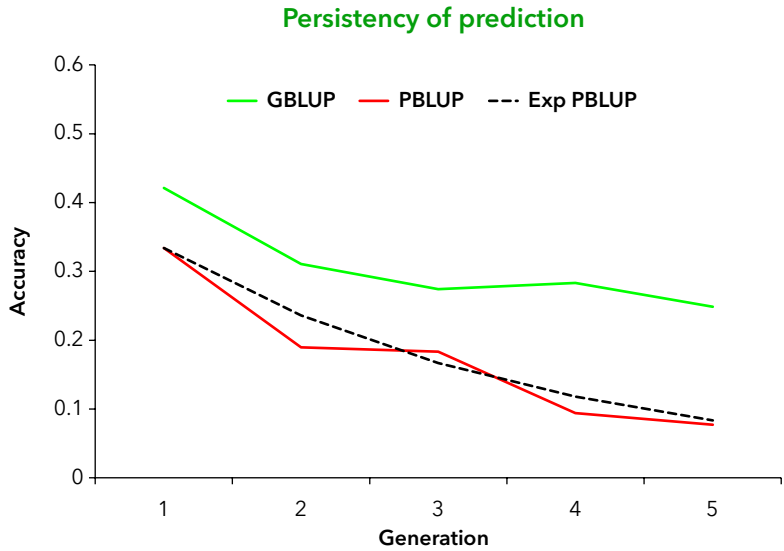

Figure 2 Persistency of accuracy of prediction for pedigree- versus genomic-based estimated breeding values with distance from the training population. Pedigree-based estimated breeding values (PBLUP) show accuracies that are expected to decline at the rate of square root of half per generation (Exp-PBLUP). Genomic-based estimated breeding values (GBLUP) show observed accuracies of greater magnitude than PBLUP. Source: modified from Wolc et al. (2011a).

van der Waaij, 2015). Initial simulations suggest that an initial training panel is sufficient for accurate selection for all subsequent generations (Meuwissen et al., 2001; Solberg et al., 2009), but practical experience shows that accuracy significantly declines as the generation distance between the training panel and the individuals to be selected increases (Wolc et al., 2011a; Fig. 2). Continual retraining (i.e. utilization of phenotypic information from the most recent generations) is required for the most accurate prediction of breeding values used for selection in subsequent generations. For some traits, there can even be an improvement in accuracy if historical generations are removed from the training panel (Wang et al., 2013).

4.2 Number of genetic markers

Early genomic selection studies utilized SNP sets containing 42 000 (Avendaño et al., 2010; Wolc et al., 2015) or 60 000 SNPs (Groenen et al., 2011). These were developed as proprietary chips using the technology available at that time. Information on the initial set of SNPs available for inclusion on these early SNP panels was obtained from a limited number of birds and thus there

was concern as to how well they represented the variation within the wider range of commercially utilized breeds (Fulton, 2012). It was hypothesized that increasing the number of SNPs genotyped would greatly enhance the accuracy of estimated breeding values as more of the variation within the genome could be captured with SNPs that were physically closer to each other and causal mutations in the genome. The subsequent sequencing of diverse chicken populations, including multiple commercially relevant broiler and egg-layer lines, identified a wider set of 139 million SNPs (Kranis et al., 2013). This SNP set was used to develop a 600 000 SNP Affymetrix Axiom chip that was publicly available and designed to provide information on SNP variation for a wide variety of lines. This 600K SNP chip was initially used for genomic selection; however, due to the nature of its development for universal use, the actual number of SNPs on this chip that were informative within any one line could be quite variable. Furthermore, many of these SNPs were in very high linkage disequilibrium (LD) and thus did not provide much additional information. Thus, while this SNP chip contained 600 000 SNPs, the actual number of SNPs that provided information for genomic selection was considerably less and very line dependent (Drobik-Czwarno et al., 2018a,b; Romé et al., 2015; Li et al., 2018; Wolc et al., 2019). The high cost of this 600K SNP chip limited its extensive use for the large numbers of birds required for genomic selection.

Subsequent studies using simulations and masking subsets of SNPs determined that genomic selection did not require 600K SNPs, and that considerably fewer SNPs provided equivalent information (Wolc unpublished; Su et al., 2012). The development of novel technology to produce lower SNP density panels (50–65K) by the major SNP chip manufacturers (ThermoFisher Applied Biosystems and Illumina) now allows for the affordable development and subsequent use of these lower-density SNP chips for genomic selection (Fig. 3).

Figure 3 Image of Applied Biosystems Axiom® myDesign™ Arrays from Thermo Fisher Scientific Axiom 600K and 65K arrays plates.

4.3 Reducing costs of genomic selection

The cost of high-density genotyping can be a limiting factor in the application of genomic selection, particularly when the number of individuals to be genotyped is high and the value of each individual is low, as is the case for chickens. Several strategies have been proposed to reduce genotyping costs.

One option is to utilize a subset of SNPs that are the best for prediction of breeding values for a specific trait (Habier et al., 2009). For highly polygenic traits, it may be difficult to identify a limited number of SNPs with good prediction accuracy. Weigel et al. (2010) suggest that at least 1500 equally spaced SNPs would be needed for genomic selection. Different traits may have few SNPs in common (Moser et al., 2010); thus, multiple trait selection would require different SNP subsets for each trait, subsequently increasing the number of SNPs required on the panel. It is quite possible that different SNPs are predictive for different breeds or populations. All of these limitations can be overcome with larger SNP panels, which could negate the cost advantage of low-density SNP panels.

Alternatively, low-density SNP sets that cover the genome at a wider spacing could be utilized, with imputation to higher density SNP. Imputation is the process of inferring missing genotypes using available information from relatives and the candidate. This is possible due to LD between SNP that are close to each other on the genome. Furthermore, the low rate of recombination between SNPs in each generation results in very tight LD between SNPs when passed from parent to progeny.

Thus, there are two major sources of information for imputation; population-wide LD and within-family segregation of haplotypes, each with their advantages and disadvantages. Modern algorithms attempt to use both of these sources. Population-wide LD utilizes information from SNPs that are in close physical proximity on the genome (Huang et al., 2009). It identifies founder haplotypes (specific combinations of SNP alleles that are transferred as a block across generations) and traces their transmission to the selection candidates. There are numerous software programs that have been developed to define these haplotypes, including AlphaImpute, Beagle, FImpute, PHASE, and fastPHASE. An advantage of this approach is that it does not require pedigree information to impute the missing SNP genotypes. However, it works best with a limited number of founders, each with multiple progeny in the population (Browning and Browning, 2009; Scheet and Stephens, 2006). Within-family segregation utilizes information from direct relatives that have high-density SNP information. If parents have high-density SNP information, the progeny genotypes can be inferred from the low-density SNP information to provide the missing high-density SNP haplotypes, using the rules of Mendelian inheritance (Habier et al., 2009; Piyasatian et al., 2010).

Genotypes of SNPs not included in the low-density panel can then be imputed using information from the parents which have high-density SNP information (Habier et al., 2009). The more costly high-density genotyping would be done on the much smaller number of parents, and that information would allow imputation of the selection candidate progeny based on the lower-cost low-density SNP genotypes. These low-density SNP sets could be designed based on either equal distance between the SNPs or based on the LD found between these SNPs. If SNPs are in LD, then limited information is obtained from any additional SNPs that are in high LD. These low-density SNP sets would likely need to be breed or even line-specific to obtain the maximum beneficial information. Various approaches to select a more informative subset of SNP have been proposed (Chang et al., 2018, 2019). One problem with these low-density panels and subsequent imputation to high-density panels is that a high information value is placed on each SNP within the low-density panel. It is assumed that the genotype information provided from the SNP chips is 100% accurate, but this is a false assumption as low-level genotyping errors can occur (Wolc and Fulton, unpublished). Any imputing or genotyping error within the low-density panel data would have a cumulative error after multiple generations of imputing. Any SNP that fail to provide results would further reduce accuracy of imputation. It is recommended that the parents of each generation be genotyped with the high-density panel (Piyasatian et al., 2010; Wang et al., 2013) to correct any genotyping errors in each generation. Improved accuracy of genomic prediction is obtained when high-density genotypes of the most influential animals are routinely generated (Judge et al., 2017).

Despite these possible risks, genomic estimated breeding values from individuals with high-density SNP information versus genomic estimated breeding values imputed from low-density SNP panel have very high correlations, ranging from 0.95 to 0.98 across various traits when both parents were high-density genotyped, validating the application of low- to high-density imputation for accurate prediction of breeding values (Wang et al., 2013). Even low-density panels of less than 1000 SNPs were shown to provide a high degree of accuracy for prediction (Wang et al., 2013; Piyasatian et al., 2010). The accuracy depends on the minor allele frequency (MAF) of SNPs, with low MAF resulting in lower accuracy of imputation (Heidaritabar et al., 2015). It is also possible to impute from medium-density panels to whole genome sequence data, but no significant gains in predictive accuracy were shown (Heidaritabar et al., 2016; Ni et al., 2017). Higher accuracy of genomically estimated breeding values is obtained if more individuals are sequenced and they are strategically selected, but the high cost of sequencing versus high-density SNP panels may not be worth this increased accuracy (Ye et al., 2018).

5 Analysis methods for genomic selection

Multiple methods have been proposed to estimate genomic breeding values from SNP data. Each has their advantages and disadvantages. There are two main concepts: Genomic BLUP (GBLUP) and its extension called single-step GBLUP, which relies on genomic relationships, and Bayesian models, which are based on estimation of SNP effects. GBLUP replaces the pedigree information with a genomic matrix relationship based on the SNP information. With only pedigree information, all individuals without own phenotype or progeny phenotype within a family have identical estimated breeding values. Utilizing the genomic matrix information allows the differentiation of superior individuals within the same family. The breeding value of an individual equals to half of the breeding value of its sire plus half of the breeding value of its dam plus Mendelian sampling. Mendelian sampling refers to a random sampling of alleles from parents. The use of genomic information for breeding value prediction exploits Mendelian sampling information in addition to the parental average information. In GBLUP, only information from individuals with genotypes is used, which often can ignore the vast majority of the phenotypic data from relatives. This limitation can be overcome by utilizing the single-step method, which combines information from all animals with phenotypes, pedigree information, and/or genotypes (Legarra et al., 2009; Christensen and Lund, 2010; Misztal et al., 2013). The single-step method enhances the accuracy of prediction not only for the genotyped individuals, but also for their non-genotyped relatives (Yan et al., 2018). The GBLUP method assumes that all SNPs have an effect on traits and those effects come from a normal distribution with the same variance. The Bayesian models relax that assumption, allowing the variation to be different for each SNP. This is the basis for both BayesA and BayesB (Meuwissen et al., 2001). Furthermore, BayesB allows for some SNPs to have no effect. The proportion of SNPs with zero effect can be estimated using the method BayesCpi. There are numerous related methods based on different prior assumptions (Habier et al., 2011; Gianola, 2013). Bayesian models were also generalized to a single-step approach to utilize information also from non-genotyped individuals (Fernando et al., 2016; Zhou et al., 2018; Gianola and Fernando, 2019).

5.1 Examples of genomic selection

All of the major poultry breeding companies (broiler and layer) are intensively using genomic selection to increase the rate of progress for multiple production traits (O'Keefe, 2009). However, the specific details of how this is being done is proprietary, with very little being published by the breeding companies. Limited publications, many utilizing simulations with population structures relevant to

commercial poultry populations, show that various methods are being explored to obtain improved accuracies in prediction (Calus et al., 2009; Piyasatian et al., 2010; Herry et al., 2018; Chu et al., 2018) and alternative implementations in breeding programs (Sitzenstock et al., 2013; Wolc et al., 2016). Which of these methods are currently being utilized for commercial poultry breeding has not been published.

Genomic predication across populations is very poor. Adding information from multiple populations increases accuracy only if the populations are closely related (Calus et al., 2014; Huang et al., 2014). Thus, genomic prediction needs to be done separately for each population. This includes the initial development and analysis of the training or reference panel.

Genomic prediction has been utilized to improve accuracy of prediction for several difficult-to-measure performance traits relevant to improved animal welfare. Any improvement in accuracy for these types of traits is helpful. Improved prediction (accuracy increased by 30%) and response to selection (by 90%) for reduced mortality following cannibalism in non-beak-treated hens has been reported (Alemu et al., 2016; Brinker et al., 2016, 2018). Genomic prediction of resistance to Marek's disease was shown to be better than pedigree selection alone (Wolc et al., 2013c). Studies with survivors of Avian Influenza showed that genomic information had predictive ability of the outcome of this disease (Wolc et al., 2018).

5.2 Prediction of Heterosis

Commercial production birds are the result of multiple line crosses, which takes advantage of heterosis caused by crossing of pure lines. Genotype information has been shown to predict heterotic effects between lines, even before the specific crosses are created (Amuzu-Aweh et al., 2013, 2015). This allows for the creation and evaluation of novel line crosses *in silico* to identify the best line combinations. These could then be confirmed *in vivo* by making those specific crosses identified as superior, which could then be tested in the commercial environment. This would not only increase the number of possible line combinations to test and reduce the time needed to test these combinations but would also significantly reduce the cost associated with testing these different line combinations in the field.

5.3 Genotype by Environment Interaction

One of the challenges in poultry breeding programs is that the environment within which the elite selection lines are maintained is very different from the environment within which the commercial production birds are housed. The elite lines are kept in optimized environments, with minimal disease

exposure, low housing density, and adequate and consistent nutrition. This well-maintained and consistent environment allows for the maximal expression of genetic potential. In contrast, the commercial production environment can be quite variable, with different climatic conditions, housing types, disease challenges, and nutrition levels. This situation can lead to a phenomenon known as genotype-by-environment interaction, where the progeny ranking can be different depending on the environment within which the phenotypes are obtained (Duenk et al., 2019). Furthermore, due to the multiplication stages between the elite lines and final commercial birds, pedigree links between the two types of birds are not available. This can be addressed by testing pedigreed progeny of sires from the elite lines in multiple commercial environments (field test or sib test). Genomics can offer additional information, as it can capture genetic similarity between animals even when pedigree information is not available, thus enabling information collected from the commercial production environment to be used in pure line selection (Chu et al., 2018).

Implementation of genomic selection in a commercial egg production line was reported by Wolc et al. (2015). In an experiment utilizing parallel genomic- and pedigree-based selection from the same base population, most of the improvements due to genomic selection suggested by prior simulations were confirmed, in particular improved accuracy of estimated breeding values, and the possibility to shorten the generation interval, resulting in a higher response to selection. Some risks were also identified, including increased inbreeding rate if proper effective population size is not maintained, undesired selection of early maturing birds if generation interval is halved in females, and the need for detailed quality control for both genomic and phenotypic data.

6 Conclusion

Genomic selection is a valuable tool, and when used in conjunction with well-established genetic selection methods can greatly improve the efficiency and accuracy of breeding programs. Genomic selection is particularly useful for the sex-limited traits of poultry egg production. The recent improvements in genetic variation identification, specifically SNP, and the lowered cost of SNP genotyping has allowed this technology to be rapidly adapted by poultry breeding companies.

7 Future trends

Genomic selection is currently done utilizing information from SNP chips, as this is currently the least costly method available for surveying extensive genomic variation. However, this surveys only those variants contained on the chip, and thus can miss novel variation found in breeds not utilized for the development

of the SNP chip. Furthermore, variation occurring in genomic regions in between these SNPs is also not surveyed, potentially missing other favorable variants. The acquisition of whole genome sequences from multiple other breeds of poultry would identify additional SNPs that could impact multiple performance traits. There are also multiple other sources of variation within the chicken genome that could impact production traits. These include copy number variants (CNV), in which small pieces of the genome are duplicated or deleted, alternative splicing, which can impact gene expression and regulation, and the presence of endogenous retroviral inserts, which can impact nearby genes (Mason et al., 2018). Recent reports have identified CNV based on SNP chip information (Drobik-Czwarno et al., 2018c). Novel technologies are needed to rapidly, accurately, and inexpensively detect these additional types of genome variation, so that this information can also be routinely used within a commercial breeding program. Additional sources of variation that can impact phenotype, such as epigenetics and metagenomics, are being studied for potential application in prediction of future performance.

In a predictive essay in 2013, Hickey (2013) proposed the following evolution of genomic selection in animal breeding. The original genomic selection model (GS 0.0) relied on LD between markers and the genome variants that influenced the specific traits (causative mutations). This was initially defined as marker-assisted selection, in which specific markers were selected that have impact on the traits of interest. The next level of genomic selection (GS 1.0) added information from relationships between close relatives (linkage) to overcome the limitations of relying solely on LD. The final level of genomic selection (GS 2.0) was envisioned as including information from genome sequences of both the founder animals and the close relatives of the selection candidates. Sequence costs are being driven down by new technologies. Low-level sequence coverage, identification of haplotypes in multiple progeny, and subsequent imputation theoretically would provide the ability to combine information from both LD and causative mutations to obtain highly accurate predictions of breeding values.

8 Where to look for further information

New research on genomic selection is published in scientific journals and presented in multiple conferences. Specific topics of interest can be searched through PubMed Central which is a database of scientific literature: https://www.ncbi.nlm.nih.gov/pmc/.

Animal Frontiers Volume 6 from 2016: https://academic.oup.com/af/issue/6/1 provides a good overview of the application of genomic selection to the improvement of livestock breeding programs.

9 References

Alemu, S. W., Calus, M. P. L., Muir, W. M., Peeters, K., Vereijken, A. and Bijma, P. 2016. Genomic prediction of survival time in a population of brown laying hens showing cannibalistic behavior. *Genet. Sel. Evol.* 48(1), 68. doi:10.1186/s12711-016-0247-4.

Amuzu-Aweh, E., Bijma, P., Kinghorn, B. P., Vereijken, A., Visscher, J., Arendonk, J. and Bovenhuis, H. 2013. Genomic prediction based on data from three layer lines using non-linear regression models. *Gen. Sel. Evol.* 46, 75. doi:10.1038/hdy.2013.7.

Amuzu-Aweh, E. N., Bovenhuis, H., de Koning, D. J. and Bijma, P. 2015. Predicting heterosis for egg production traits in crossbred offspring of individual White Leghorn sires using genome-wide SNP data. *Genet. Sel. Evol.* 47, 27. doi:10.1186/s12711-015-0088-6.

Avendaño, S., Watson, K. A. and Kranis, A. 2010. Genomics in poultry breeding: from Utopia to deliverables. Session 07-01. *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*, 1–6 August 2010, Leipzig, Germany.

Brinker, T., Raymond, B., Bijma, P., Vereijken, A. and Ellen, E. D. 2016. Estimation of total genetic effects for survival time in crossbred laying hens showing cannibalism, using pedigree or genomic information. *J. Anim. Breed. Genet.* 134, 60–8.

Brinker, T., Bijma, P., Vereijken, A. and Ellen, E. D. 2018. The genetic architecture of socially-affected traits: a GWAS for direct and indirect genetic effects on survival time in laying hens showing cannibalism. *Genet. Sel. Evol.* 50, 38. doi:10.1186/s12711-018-0409-7.

Browning, B. L. and Browning, S. R. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* 84(2), 210–23. doi:10.1016/j.ajhg.2009.01.005.

Calus, M. P. L., Meuwissen, T. H. E., Windig, J. J., Knol, E. F., Schrooten, C., Vereijken, A. L. J. and Veerkamp, R. F. 2009. Effects of the number of markers per haplotype and clustering of haplotypes on the accuracy of QTL mapping and prediction of genomic breeding values. *Genet. Sel. Evol.* 41, 11. doi:10.1186/1297-9686-41-11.

Calus, M. P. L., Huang, H., Vereijken, A., Visscher, J., ten Napel, J. and Windig, J. J. 2014. Genomic prediction based on data from three layer lines: a comparison between linear methods. *Genet. Sel. Evol.* 46, 57. doi:10.1186/s12711-014-0057-5.

Chang, L. Y., Toghiani, S., Ling, A., Aggrey, S. E. and Rekaya, R. 2018. High density marker panels, SNPs prioritizing and accuracy of genomic selection. *BMC Genet.* 19(1), 4. doi:10.1186/s12863-017-0595-2.

Chang, L. Y., Toghiani, S., Aggrey, S. E. and Rekaya, R. 2019. Increasing accuracy of genomic selection in presence of high density marker panels through the prioritization of relevant polymorphisms. *BMC Genet.* 20(1), 21. doi:10.1186/s12863-019-0720-5.

Christensen, O. F. and Lund, M. S. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42, 2. doi:10.1186/1297-9686-42-2.

Chu, T. T., Alemu, S. W., Norberg, E., Sørensen, A. C., Henshall, J., Hawken, R. and Jensen, J. 2018. Benefits of testing in both bio-secure and production environments in genomic selection breeding programs for commercial broiler chicken. *Genet. Sel. Evol.* 50, 52. doi:10.1186/s12711-018-0430-x.

Drobik-Czwarno, W., Wolc, A., Fulton, J. E., Jankowski, T., Arango, J., O'Sullivan, N. P. and Dekkers, J. C. M. 2018a. Genetic basis of resistance to avian influenza in different commercial varieties of layer chickens. *Poult. Sci.* 97(10), 3421–8. doi:10.3382/ps/pey233.

Drobik-Czwarno, W., Wolc, A., Fulton, J. E., Arango, J., Jankowski, T., Settar, P., O'Sullivan, N. P. and Dekkers, J. C. M. 2018b. Identifying genetic basis for resistance to avian influenza in commercial egg layer chickens. *Animal* 12(7), 1363–71.

Drobik-Czwarno, W., Wolc, A., Fulton, J. E. and Dekkers, J. C. M. 2018c. Detection of copy number variations in brown and white layers based on genotyping panels with different densities. *Genet. Sel. Evol.* 50(1), 54. doi:10.1186/s12711-018-0428-4.

Duenk, P., Calus, M. P. L., Wientjes, Y. C. J., Breen, V. P., Henshall, J. M., Hawken, R. H. and Bijma, P. 2019. Estimating the purebred-crossbred genetic correlation of body weight in broiler chickens with pedigree or genomic relationships. *Genet. Sel. Evol.* 51, 6. doi:10.1186/s12711-019-0447-9.

Fernando, R. L., Cheng, H., Golden, B. L. and Garrick, D. J. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. *Genet. Sel. Evol.* 48(1), 96. doi:10.1186/s12711-016-0273-2.

Fulton, J. E. 2012. Genomic selection for poultry breeding. *Anim. Front.* 2(1), 30–6. doi:10.2527/af.2011-0028.

Gianola, D. 2013. Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* 194(3), 573–96. doi:10.1534/genetics.113.151753.

Gianola, D. and Fernando, R. L. 2019. A multiple-trait Bayesian Lasso for genome-enabled analysis and presiction of complex traits. *bioRziv* 852749. doi:10.1101/852749.

Groenen, M. A., Megens, H.-J., Zare, Y., Warren, W. C., Hillier, L. W., Crooijmans, R. P., Vereijken, A., Okimoto, R., Muir, W. M. and Cheng, H. H. 2011. The development and characterization of a 60K SNP chip for chicken. *BMC Genomics* 12(1), 274. doi:10.1186/1471-2164-12-274.

Habier, D., Fernando, R. L. and Dekkers, J. C. M. 2009. Genomic selection using low-density marker panels. *Genetics* 182(1), 343–53. doi:10.1534/genetics.108.100289.

Habier, D., Fernando, R. L., Kizilkaya, K. and Garrick, D. J. 2011. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12, 186. doi:10.1186/1471-2105-12-186.

Heidaritabar, M., Calus, M. P. L., Vereijken, A., Groenen, M. A. M. and Bastiaansen, J. W. M. 2015. Accuracy of imputation using the most common sires as reference population in layer chickens. *BMC Genet.* 16, 101. doi:10.1186/s12863-015-0253-5.

Heidaritabar, M., Calus, M. P. L., Megens, H. J., Vereijken, A., Groenen, M. A. M. and Bastiaansen, J. W. M. 2016. Accuracy of genomic prediction using imputed whole-genome sequence data in white layers. *J. Anim. Breed. Genet.* 133(3), 167–79. doi:10.1111/jbg.12199.

Herry, F., Herault, F., Druet, D. P., Varenne, A., Burlot, T., Le Roy, P. and Allais, S. 2018. Design of low density SNP chips for genotype imputation in layer chicken. *BMC Genomics* 19, 108. doi:10.1186/s12863-018-0695-7.

Hickey, J. M. 2013. Sequencing millions of animals for genomic selection 2.0. *J. Anim. Breed. Genet.* 130(5), 331–2. doi:10.1111/jbg.12054.

Hillier, L. W, Miller, W., Birney, E., Warren, W., Hardison, R. C., International Chicken Genome Sequencing Consortium, et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432(7018), 695–716. doi:10.1038/nature03154.

Huang, L., Wang, C. and Rosenburg, N. A. 2009. The relationship between imputation error and statistical power in genetic association studies in diverse populations. *Am. J. Hum. Genet.* 85(5), 692–8. doi:10.1016/j.ajhg.2009.09.017.

Huang, H., Windig, J. J., Vereijken, A. and Calus, M. P. L. 2014. Genomic prediction based on data from three layer lines using non-linear regression models. *Genet. Sel. Evol.* 46, 75. doi:10.1186/s12711-014-0075-3.

Judge, M. M., Purfield, D. C., Sleator, R. D. and Berry, D. P. 2017. The impact of multi-generational genotype imputation strategies on imputation accuracy and subsequent genomic predictions. *J. Anim. Sci.* 95(4), 1489–501. doi:10.2527/jas.2016.1212.

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., Talbot, R., Pirani, A., Brew, F., Kaiser, P., Hocking, P. M., Fife, M., Salmon, N., Fulton, J., Strom, T. M., Haberer, G., Weigend, S., Preisinger, R., Gholami, M., Qanbari, S., Simianer, H., Watson, K. A., Woolliams, J. A. and Burt, D. W. 2013. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14, 59. doi:10.1186/1471-2164-14-59.

Legarra, A., Aguilar, I. and Misztal, I. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92(9), 4656–63. doi:10.3168/jds.2009-2061.

Li, S., Wang, X., Qu, L., Dou, T., Ma, M., Shen, M., Guo, J., Hu, Y. and Wang, K. 2018. Genome-wide association studies for small intestine length in an F2 population of chickens. *Ital. J. Anim. Sci.* 17(2), 294–300. doi:10.1080/1828051X.2017.1368419.

Lush, J. L. 1943. *Animal Breeding Plans*. Iowa State College Press, Ames, IA.

Mason, A. S., Wolc, A., Arango, J., Settar, P., Lund, A. R., Burt, D. W. and Fulton, J. E. 2018. Effect of endogenous retroviral elements (ALVE) on egg size in commercial egg production lines. *Proceedings of the World Congress on Genetics Applied to Livestock Production*, WCGALP 11.59.

Meuwissen, T. H. E., Hayes, B. J. and Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157(4), 1819–29.

Misztal, I., Aggrey, S. E. and Muir, W. M. 2013. Experiences with a single-step genome evaluation. *Poult. Sci.* 92(9), 2530–4. doi:10.3382/ps.2012-02739.

Moser, G., Khatkar, M. S., Hayes, B. J. and Raadsma, H. W. 2010. Accuracy of direct genomic values in Holstein bulls and cows using subsets of SNP markers. *Genet. Sel. Evol.* 42, 37. doi:10.1186/1297-9686-42-37.

Muir, W. M., Wong, G. K., Zhang, Y., Wang, J., Groenen, M. A. M., Crooijmans, R. P. M. A., Megens, H. -J., Zhang, H. M., Mckay, J. C., McLeod, S., Okimoto, R., Fulton, J. E., Settar, P., O'Sullivan, N. P., Vereijken, A., Jungerius-Rattink, A., Albers, G. A. A., Taylor Lawley, C., Delany, M. E. and Cheng, H. H. 2008. Review of the initial validation and characterization of a 3K chicken SNP array. *Worlds Poult. Sci. J.* 64(2), 219–26. doi:10.1017/S0043933908000019.

Ni, G., Cavero, D., Fangmann, A., Erbe, M. and Simianer, H. 2017. Whole-genome sequence-based genomic prediction in laying chickens with different genomic relationship matrices to account for genetic architecture. *Genet. Sel. Evol.* 49(1), 8. doi:10.1186/s12711-016-0277-y.

O'Keefe, T. 2009. Genomic selection: the race is on. *WATT Global Media*. Available at: https://www.wattagnet.com/articles/3894-genomic-selection-the-race-is-on.

Oldenbroek, K. and van der Waaij, L. 2015. Textbook Animal Breeding and Genetics for BSc students. Centre for Genetic Resources The Netherlands and Animal Breeding and Genomics Centre, 2015. *Groen Kennisnet*. Available at: https://wiki.groenkennisnet.nl/display/TAB/.

Piyasatian, N., Habier, D., Kranis, A., Watson, K. A., Avendano, S., Fernando, R. L. and Dekkers, J. C. M. 2010. Genomic selection using evenly spaced low-density SNPs. *World Congress Applied to Livestock Production*.

Pszczola, M., Strabel, T., Mulder, H. A. and Calus, M. P. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. *J. Dairy Sci.* 95(1), 389–400. doi:10.3168/jds.2011-4338.

Romé, H., Varenne, A., Hérault, F., Chapuis, H., Alleno, C., Dehais, P., Vignal, A., Burlot, T. and Le Roy, P. 2015. GWAS analyses reveal QTL in egg layers that differ in response to diet differences. *Genet. Sel. Evol.* 47, 83. doi:10.1186/s12711-015-0160-2.

Scheet, P. and Stephens, M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78(4), 629–44. doi:10.1086/502802.

Sitzenstock, F., Ytournel, F., Sharifi, A. R., Cavero, D., Täubert, H., Preisinger, R. and Simianer, H. 2013. Efficiency of genomic selection in an established commercial layer breeding program. *Genet. Sel. Evol.* 45(1), 29. doi:10.1186/1297-9686-45-29.

Solberg, T. R., Sonesson, A. K., Woolliams, J. A., Ødegard, J. and Meuwissen, T. H. E. 2009. Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. *Genet. Sel. Evol.* 41, 53. doi:10.1186/1297-9686-41-53.

Su, G., Brøndum, R. F., Ma, P., Guldbrandtsen, B., Aamand, G. P. and Lund, M. S. 2012. Comparison of genomic predictions using medium-density (~54,000) and high-density (~777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *J. Dairy Sci.* 95(8), 4657–65. doi:10.3168/jds.2012-5379.

Wang, C., Habier, D., Peiris, B. L., Wolc, A., Kranis, A., Watson, K. A., Avendano, S., Garrick, D. J., Fernando, R. L., Lamont, S. J. and Dekkers, J. C. 2013. Accuracy of genomic prediction using an evenly spaced, low-density single nucleotide polymorphism panel in broiler chickens. *Poult. Sci.* 92(7), 1712–23. doi:10.3382/ps.2012-02941.

Weigel, K. A., de los Campos, G., Vasquez, A. I., Rosa, G. J. M., Gianola, D. and Van Tassell, C. P. 2010. Accuracy of direct genomic values derived from imputed single nucleotide polymorphism genotypes in Jersey cattle. *J. Dairy Sci.* 93(11), 5423–35. doi:10.3168/jds.2010-3149.

Wolc, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Preisinger, R., Habier, D., Fernando, R., Garrick, D. J., Lamont, S. J. and Dekkers, J. C. M. 2011a. Persistence of accuracy of estimated breeding values in layers using marker and pedigree based relationship matrices. *Genet. Sel. Evol.* 43, 23.

Wolc, A., Stricker, C., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N., Habier, D., Fernando, R., Garrick, D. J., Lamont, S. J. and Dekkers, J. C. M. 2011b. Breeding value prediction for production traits in layers using pedigree and marker based methods. *Genet. Sel. Evol.* 43, 5.

Wolc, A., Arango, J., Jankowski, T., Settar, P., Fulton, J. E., O'Sullivan, N. P., Fernando, R., Garrick, D. J. and Dekkers, J. C. M. 2013a. Pedigree and genomic analyses of feed consumption and residual feed intake in laying hens. *Poult. Sci.* 92, 2270–5.

Wolc, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Preisinger, R., Fernando, R., Garrick, D. J. and Dekkers, J. C. M. 2013b. Analysis of egg production in layer chickens using a random regression model with genomic relationships. *Poult. Sci.* 92(6), 1486–91. doi:10.3382/ps.2012-02882.

Wolc, A., Arango, J., Jankowski, T., Settar, P., Fulton, J. E., O'Sullivan, N. P., Fernando, R., Garrick, D. J. and Dekkers, J. C. M. 2013c. Genome-wide association study for Marek's disease mortality in layer chickens. *Avian Dis.* 57(2 Suppl.), 395–400. doi:10.1637/10409-100312-Reg.1.

Wolc, A., Zhao, H. H., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Preisinger, R., Stricker, C., Habier, D., Fernando, R. L., Garrick, D. J., Lamont, S. J. and Dekkers, J. C. M. 2015. Response and inbreeding from a genomic selection experiment in layer chickens. *Genet. Sel. Evol.* 47, 59. doi:10.1186/s12711-015-0133-5.

Wolc, A., Kranis, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P., Avendano, A., Watson, K. A., Hickey, J. M., de los Campos, G., Fernando, R. L., Garrick, D. J. and Dekkers, J. C. M. 2016. Implementation of genomic selection in the poultry industry. *Anim. Front.* 6(1), 23–31. doi:10.2527/af.2016-0004.

Wolc, A., Drobik-Czwarno, W., Fulton, J. E., Arango, J., Jankowski, T. and Dekkers, J. C. M. 2018. Genomic prediction for avian influenza infection outcome in layer chickens. *Genet. Sel. Evol.* 50(1), 21. doi:10.1186/s12711-018-0393-y.

Wolc, A., Arango, J., Settar, P., Fulton, J. E., O'Sullivan, N. P. and Dekkers, J. C. M. 2019. Genetics of male reproductive performance in White Leghorns. *Poult. Sci.* 98(7), 2729–33. doi:10.3382/ps/pez077.

Wong, G. K., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z., Meng, Q., Zhou, J., Li, D., Zhang, J., Ni, P., Li, S., Ran, L., Li, H., Zhang, J., Li, R., Li, S., Zheng, H., Lin, W., Li, G., Wang, X., Zhao, W., Li, J., Ye, C., Dai, M., Ruan, J., Zhou, Y., Li, Y., He, X., Zhang, Y., Wang, J., Huang, X., Tong, W., Chen, J., Ye, J., Chen, C., Wei, N., Li, G., Dong, L., Lan, F., Sun, Y., Zhang, Z., Yang, Z., Yu, Y., Huang, Y., He, D., Xi, Y., Wei, D., Qi, Q., Li, W., Shi, J., Wang, M., Xie, F., Wang, J., Zhang, X., Wang, P., Zhao, Y., Li, N., Yang, N., Dong, W., Hu, S., Zeng, C., Zheng, W., Hao, B., Hillier, L. W., Yang, S. P., Warren, W. C., Wilson, R. K., Brandström, M., Ellegren, H., Crooijmans, R. P., van der Poel, J. J., Bovenhuis, H., Groenen, M. A., Ovcharenko, I., Gordon, L., Stubbs, L., Lucas, S., Glavina, T., Aerts, A., Kaiser, P., Rothwell, L., Young, J. R., Rogers, S., Walker, B. A., van Hateren, A., Kaufman, J., Bumstead, N., Lamont, S. J., Zhou, H., Hocking, P. M., Morrice, D., de Koning, D. J., Law, A., Bartley, N., Burt, D. W., Hunt, H., Cheng, H. H., Gunnarsson, U., Wahlberg, P., Andersson, L., Kindlund, E., Tammi, M. T., Andersson, B., Webber, C., Ponting, C. P., Overton, I. M., Boardman, P. E., Tang, H., Hubbard, S. J., Wilson, S. A., Yu, J., Wang, J., Yang, H. and International Chicken Polymorphism Map Consortium. 2004. A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms. *Nature* 432(7018), 717–22. doi:10.1038/nature03156.

Yan, Y., Wu, G., Liu, A., Sun, C., Han, W., Li, G. and Yang, N. 2018. Genomic prediction in a nuclear population of layers using single-step models. *Poult. Sci.* 97(2), 397–402. doi:10.3382/ps/pex320.

Ye, S., Yuan, X., Lin, X., Gao, N., Luo, Y., Chen, Z., Li, J., Zhang, X. and Zhang, Z. 2018. Imputation from SNP chip to sequence: a case study in a Chinese indigenous chicken population. *J. Anim. Sci. Biotech.* 9, 30. doi:10.1186/s40104-018-0241-5.

Zhou, L., Mrode, R., Zhang, S., Zhang, Q., Li, B. and Liu, J. F. 2018. Factors affecting GEBV accuracy with single-step Bayesian models. *Heredity (Edinb)* 120(2), 100–9. doi:10.1038/s41437-017-0010-9.